Fault Diagnosis Based on Interpretable Convolutional Temporal-spatial Attention Network for Offshore Wind Turbines

Xiangjing Su, Chao Deng, Yanhao Shan, Farhad Shahnia, Yang Fu, and Zhaoyang Dong

***Abstract*—Fault diagnosis (FD) for offshore wind turbines (WTs) are instrumental to their operation and maintenance (O&M). To improve the FD effect in the very early stage, a condition monitoring based sample set mining method from supervisory control and data acquisition (SCADA) time-series data is proposed. Then, based on the convolutional neural network (CNN) and attention mechanism, an interpretable convolutional temporal-spatial attention network (CTSAN) model is proposed. The proposed CTSAN model can extract deep temporal-spatial features from SCADA time-series data sequentially by: ① a convolution feature extraction module to extract features based on time intervals; ② a spatial attention module to extract spatial features considering the weights of different features; and ③ a temporal attention module to extract temporal features considering the weights of intervals. The proposed CTSAN model has the superiority of interpretability by exposing the deep temporal-spatial features extracted in a human-understandable form of the temporal-spatial attention weights. The effectiveness and superiority of the proposed CTSAN model are verified by real offshore wind farms in China.**

***Index Terms*—Offshore wind turbine (WT), gearbox, fault diagnosis (FD), attention mechanism, interpretability, temporal-spatial feature.**

I. Introduction

WIND energy has received worldwide attention due to its cleanness and sustainability, and plays an important role in the energy market [1]. Offshore wind power is richer compared with onshore wind power [2]. However, offshore wind turbines (WTs) face the following specific challenges: ① harsher environment, changeable sea conditions, and the developing trend of larger-scale units lead to a higher failure rate of offshore WTs; and ② due to the limitation of meteorological and transportation conditions, the accessibility of offshore WTs is poor, the maintenance cycle is long, and their operation and maintenance (O&M) costs can reach up to 25% of the total revenue [2].

Therefore, the fault diagnosis (FD) for offshore WTs is critical for reducing O&M costs and enhancing the overall generated power.

In recent years, scholars have conducted extensive exploration and research on the FD for offshore WTs, mainly including model-based methods [3]-[5] and data-driven methods [6]-[14]. Among them, the model-based methods establish a FD model by utilizing the correlation between sensor data and system stage of WTs. However, the diagnosis accuracy of the FD model is significantly susceptible to variations in system variables. Besides, as WTs evolve rapidly in size and functionality, the applications of quantitative model analysis have become increasingly difficult and even impracticable [15]. In contrast, the data-driven methods avoid the need for exact physical models of WTs. These methods are based on the manipulation of voluminous datasets gathered by WT sensors, the extraction of pivotal features via intelligent algorithms, and the utilization of classification functions to realize FD [16]. For example, the multi-layer perceptron (MLP) model is implemented for FD in bearings, marking one of the earliest applications of machine learning (ML) in the domain of FD for offshore WTs [6]. In [7], a sequential feature selection algorithm is utilized to ascertain predictive variables from supervisory control and data acquisition (SCADA) time-series data, subsequently facilitating multi-fault FD for WT gearboxes via support vector machines (SVMs). In [8], the Gaussian process regression is employed to extract fault features and the random forest (RF) technology is used for FD, improving the computational efficiency while maintaining the diagnosis accuracy. However, the algorithms above are fundamentally shallow ML algorithms and become incapable to handle the SCADA data of offshore wind power. On the one hand, the SCADA data of offshore wind power characterized by non-stationarity, non-linearity, and high-noise pose limitations to the diagnosis accuracy of these ML algorithms. On the other hand, the SCADA data of offshore wind power are unable to circumvent the "curse of dimen-

Manuscript received: August 27, 2023; revised: January 18, 2024; accepted: March 9, 2024. Date of CrossCheck: March 9, 2024. Date of online publication: April 5, 2024.

X. Su, C. Deng, and Y. Fu (corresponding author) are with the Engineering Research Center of Offshore Wind Technology Ministry of Education, Shanghai University of Electric Power, Shanghai 200090, China, and X. Su and Y. Fu are also with the Offshore Wind Power Research Institute, Shanghai University of Electric Power, Shanghai 200090, China (e-mail: xiangjing_su@126.com; dengchaoll@foxmail.com; mfudong@126.com).

Y. Shan is with the Yantai Power Supply Company, State Grid Shandong Electric Power Co., Ltd., Yantai 264001, China (e-mail: 954636907@qq.com).

F. Shahnia is with the School of Engineering and Energy, Murdoch University, Perth WA 6150, Australia (e-mail: f.shahnia@murdoch.edu.au).

Z. Dong is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mail: zydong@ieee.org).

DOI: 10.35833/MPCE.2023.000606

sion", i.e., the SCADA system generates a vast amount of historical data. With the escalation in data dimensions, the model computation cost increases significantly, adversely impacting the model generalization and diagnosis performance.

In contrast, due to the superiority in hierarchical structure extraction and deep feature representation, deep learning (DL) methods have gradually become the mainstream in the field of FD for WT. For instance, an FD methodology based on the generalized model-agnostic meta-learning (MAML) is proposed, which employs multi-kernel efficient channel attention to construct a channel interaction feature encoder, thereby facilitating multi-fault FD effectively [9]. To capture the variable space differences between fault and normal features, a multi-scale convolution neural network (CNN) architecture that extracts multi-scale features to facilitate FD for WT gearboxes are proposed [10]. To capture the temporal features of SCADA data during the FD process, the robust memory capabilities of long short-term memory (LSTM) networks in multi-variable time-series prediction is leveraged to identify potential WT fault types [11]. In [12], a WT FD methodology is proposed based on the temporal-spatial fusion neural network, which extracts multi-scale spatial characteristics among different variables of SCADA data through multi-convolution kernels of varying sizes, and then LSTM is employed to learn temporal features. In [13], a deep echo state network and multi-scale convolution residual network are utilized to propose a spatio-temporal multi-scale neural network (STMNN) with parallel feature extraction, achieving superior multi-type FD. In [14], the variability of WT operation state is considered and an FD method is proposed based on temporal-spatial features and neighboring operation state, reducing the incidence of FD resulting from sudden changes in WT operation.

Although DL-based FD methods have significantly enhanced the diagnosis performance by extracting temporal-spatial features, there are still some challenges. Firstly, existing studies such as CNN-based methods in [10]-[13] can only extract local features but cannot easily achieve complete and dynamic feature extraction. When extracting temporal correlations among data, the undifferentiated information compression in the hidden layers weakens the temporal differences of the input information, failing to highlight the effects of key historical information. Secondly, the fault development of WTs is commonly a gradual process of normal state to early fault to fault [17], [18]. That is, early fault may appear earlier than the moment of being detected by the SCADA system. Thus, existing FD studies based on SCADA data are incapable of mining early fault in sample datasets and making the warning early before the fault happens. Most importantly, existing DL-based FD methods are often referred to as black box, lacking interpretability and the ability of presenting temporal-spatial correlations extracted from SCADA data in a human-understandable format [19]. Consequently, it is challenging to ascertain whether the model has captured the key features, hindering its deployment in practical applications [20]. So far, only a few studies have paid attention to the interpretability of FD by DL-based methods [21]. For example, the attention weights of transformer model are utilized to achieve an interpretable rotating machinery FD [22]. A rotating machinery FD is proposed to achieve the interpretability of fault types and motor feature signals through attention weight parameterization [23]. To our best knowledge, the interpretable DL has not been employed in FD for offshore WTs, and no study has designed a temporal-spatial interpretable diagnosis model.

To address the challenges above, firstly, this paper designs an early fault sample set mining method from SCADA data, and then proposes an interpretable convolutional temporal-spatial attention network (CTSAN) model based on CNN and attention mechanism of early FD for offshore WTs. The feasibility and superiority of the proposed CTSAN model are verified in a real offshore wind farm in China. The main contributions of this paper are summarized as follows.

1) An early fault sample set mining method based on condition monitoring technology for SCADA time-series data is proposed, which supplies the FD model with sufficient WT early fault sample sets.

2) A CTSAN model is proposed. It can not only diagnose WT early faults accurately by capturing critical temporal-spatial correlations from the SCADA time-series data, but also be interpretable by presenting the temporal-spatial correlations in the form of corresponding attention weights.

The rest of this paper is arranged as follows. Section II introduces the proposed early fault sample set mining method. On this basis, Section III presents the interpretable FD process based on the proposed CTSAN model, while Section IV expounds the complete flowchart of the proposed FD for offshore WTs. Finally, in Section V, two real cases are used to verify the effectiveness and superiority of the proposed CTSAN model in terms of accuracy and interpretability, and the conclusions are drawn in Section VI.

II. Early Fault Sample Set Mining Method

The condition monitoring technology through the extraction of monitoring-sensitive indicators can be used to divide the WT health into normal state and early fault stage boundaries, enabling the recognition of early faults [24], [25]. This paper employs this technology, where the actual starting point of the fault is determined by establishing the normal behavior model (NBM) of the target modeling parameters and monitoring the output residual error using the exponentially weighted moving average (EWMA) control chart.

A. Description of WT Fault Process

The WT health can be divided into three stages, i.e., normal state, early fault, and fault stages [18], as shown in Fig. 1. Assume t_1 as the dividing point (i.e., the actual starting point of the fault) between the normal state and early fault stages. When the fault intensifies and an obvious fault phenomenon appears, e.g., the temperature exceeding the limit, the SCADA system detects the fault at t_2. After the fault removal, the internal stability of the WT is restored and values of all variables return to normal state at t_3.

Through the SCADA alarm log and O&M log, the prior knowledge of the WT faults can be determined to study the fault case. By querying the occurrence and disappearance

time of the specific fault code, t_2 and t_3 can be determined. Since t_1 cannot be easily determined, existing FD models mainly take t_2 as the dividing point to mine the fault sample set. If the fault sample set is mined based on the SCADA data between t_2 and t_3, it is difficult to start FD at an early stage, which not only threatens the safety of WT health, but also increases the O&M cost of WTs. Therefore, if t_1 is determined, the fault sample sets can be mined based on the data between t_1 and t_3, and the FD model can then effectively diagnose the early fault.

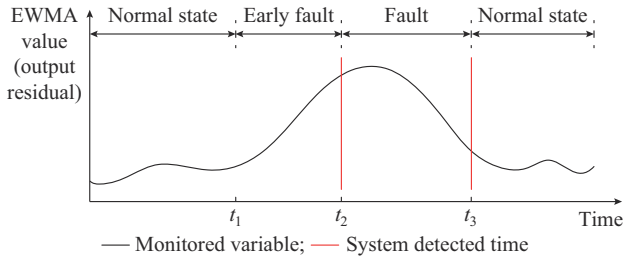

Fig. 1. Assumed process of WT health.

B. NBM

NBM is a widely used data-driven condition monitoring strategy. Based on the input parameters of the FD model, the NBM of the target modeling parameters is built according to the SCADA time-series data of normal operation, and the output values are predicted. By comparing the residual between the actual parameter value and the predicted value, the component condition monitoring is realized. This paper implements the NBM based on convolutional neural network and gated recurrent unit (CNN-GRU) method [26] and verifies it through experiments. Compared with shallow and other DL-based models, the NBM achieves better condition monitoring performance.

Since the NBM is based on SCADA time-series data, the NBM process for time-series data will be introduced first. Taking a single sample set as an example, $y^{(t)}$ is the measured value of the target modeling parameter at time t. Set $X=\{X^{(1)},X^{(2)},\ldots,X^{(T)}\}$ as the input sequence of other operating and environmental parameter data at the corresponding time, the NBM process of SCADA time-series data can be expressed as:

$$\hat{y}^{(T)}=f_\theta(X)=f_\theta(x^{(1)},x^{(2)},\ldots,x^{(T)}) \tag{1}$$

where $x^{(t)}=\{x_1^{(t)},x_2^{(t)},\ldots,x_i^{(t)},\ldots,x_m^{(t)}\}$ is the set of measured values of m input features at time t, $x_i^{(t)}$ is the measured value of the i^{th} variable at time t; $\hat{y}^{(T)}$ is the target modeling parameter at time T; $f_\theta(\cdot)$ is the model function; and $X=\{x^{(1)},x^{(2)},\ldots,x^{(T)}\}$ is the sequence of input variables from time 1 to T.

C. Monitoring Based on EWMA Control Charts

The NBM is used based on normal operational data of WT. Under the normal state stage, the residual between the actual value and predicted value of the target modeling parameter fluctuates around the zero-mean value [25], so the change of the output residual can reveal a potential fault. The output residual at time t can be calculated by:

$$R_t=y_t^{(T)}-\hat{y}_t^{(T)} \tag{2}$$

where $y_t^{(T)}$ is the actual value of target modeling parameter at time t; and $\hat{y}_t^{(T)}$ is the predicted value of target modeling parameter at time t.

To effectively filter out noises and improve the accuracy, the EWMA control chart is used to monitor the output residual sequence of the test set [27]. The EWMA value at time t can be expressed as:

$$EWMA_t=\begin{cases}\lambda R_t+(1-\lambda)\cdot EWMA_{t-1} & t>0\\ R_{\text{mean}} & t=0\end{cases} \tag{3}$$

where R_{mean} is the mean value of output residuals under all normal working conditions; and λ is the weight of historical data on current EWMA value (assumed as 0.2 in this paper [27]). The time window of EWMA control chart is the length of the test set.

Because the fault of WT is usually accompanied by the increase of the actual value of the target modeling parameter, which leads to the increase of the output residual, this paper only focuses on the upper control limit (UCL) of the EWMA control chart to monitor the gearbox and identify its abnormal stage [28].

$$UCL(t)=R_{\text{mean}}+K\sigma\sqrt{\frac{\lambda[1-(1-\lambda)^{2t}]}{2-\lambda}} \tag{4}$$

where $UCL(t)$ is the UCL value at time t; σ is the standard deviation of the output residual; and K is the threshold coefficient, which is assumed to be 3 in this paper [27].

The data sampling and labeling process of SCADA data is shown schematically in Fig. 2. As the WT has the trend of early fault, the output residual of the target modeling parameters gradually rises, leading to the rise of the EWMA value. t_1 is defined as the time that the EWMA value reaches the UCL, i.e., the time the WT changes from the normal state stage to the early fault stage, as shown in Fig. 2.

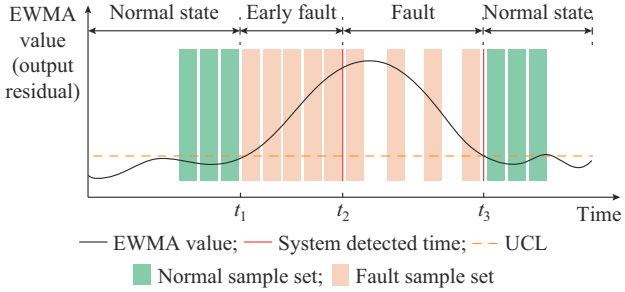

Fig. 2. Sampling and labeling process of SCADA data.

D. Data Sampling and Labeling

After determining t_1, t_2, and t_3, set the total length of the SCADA sequence as R and the number of variables as m. Now, the early fault data sequence between t_1 and t_2 can be expressed as $X_{\text{F1}}=\{x^{(t_1)},x^{(t_1+1)},\ldots,x^{(t_2)}\}\in\mathbb{R}^{m\times(t_2-t_1)}$. Likewise, the fault sequence data between t_2 and t_3 are given by $X_{\text{F2}}=\{x^{(t_2)},x^{(t_2+1)},\ldots,x^{(t_3)}\}\in\mathbb{R}^{m\times(t_3-t_2)}$ while the normal sequence data between $1-t_1$ and t_3-R are defined as $X_{\text{Normal}}=\{x^{(1)},x^{(2)},\ldots,x^{(t_1)}\}\bigcup\{x^{(t_3)},x^{(t_3+1)},\ldots,x^{(R)}\}$, where $\bigcup$ is the union function. Then, the sample set can be obtained by a sliding window. Set L, and K as the length and step size of sliding

window, respectively. The obtained N sample sets are $\{S_1, S_2, \ldots, S_N\}$, where $S_1 = \{x^{(1)}, x^{(2)}, \ldots, x^{(L)}\} \in \mathbb{R}^{L \times m}$, $S_2 = \{x^{(1+K)}, x^{(2+K)}, \ldots, x^{(L+K)}\} \in \mathbb{R}^{L \times m}$, and so on.

The error classification is most likely to occur in the early fault stage [18]. Therefore, to ensure that the FD model can effectively detect early fault, the early fault sample set of S_{F1} is made by X_{F1} at a high sampling rate (small steps) and the fault sample set of S_{F2} is made by X_{F2} at a low sampling rate (large steps). Merging the above two sample sets to form the overall fault sample sets, let $S_F = S_{F1} \bigcup S_{F2}$, the corresponding true label indicates the fault label. The normal sample set of S_{Normal} is made by X_{Normal} at a specific sampling rate and its corresponding true label is 0, which means the normal stage.

III. Interpretable FD Process Based on Proposed CTSAN Model

An interpretable CTSAN model is proposed and applied in this paper. The structure characteristics of the proposed CTSAN model enable it to extract deep temporal-spatial features from the SCADA time-series data. At the same time, the extracted temporal-spatial features can be presented in the form of temporal and spatial attention weights, directly determining the parameter or time interval that contributes most to classification, and realizing the interpretability within the model.

The proposed CTSAN model contains four sequentially connected feature extraction modules: ① convolution feature extraction module; ② spatial attention module; ③ temporal attention module; and ④ classification module.

Figure 3 shows the structure of the proposed CTSAN model. As observed from Fig. 3, the convolution feature extraction module extracts features C based on intervals from the SCADA input sequence. Then, the spatial attention module extracts the global spatial features C' of all intervals considering the spatial attention weight. Later, the global spatial features of all intervals C' are sent to the temporal attention module to extract the global temporal-spatial features C'', considering the weight of each interval. Finally, C'' is sent to the classification module to extract the predicted label corresponding to the input sequence.

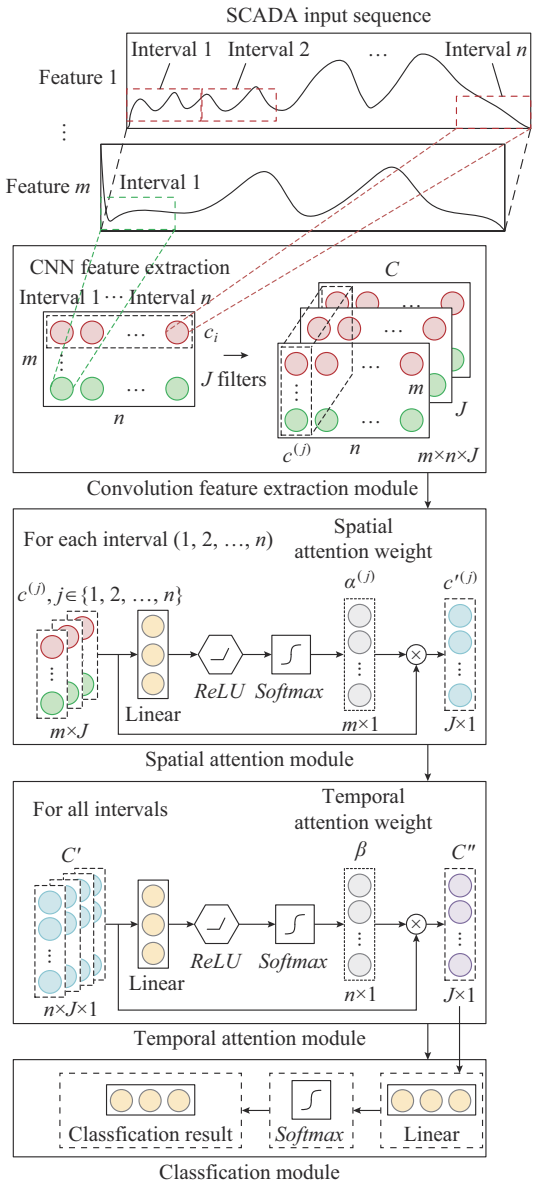

Fig. 3. Structure of proposed CTSAN model.

A. Convolution Feature Extraction Module

To reduce the computational complexity, the CNN is used to extract the key features based on time interval from the input multi-variate time series to transform the model focus from the huge number of time points to less intervals. Taking a single sample set S_1 as an example, the input sequence of multi-variate time series $X = \{x^{(1)}, x^{(2)}, \ldots, x^{(T)}\} \in \mathbb{R}^{m \times T}$ with a length of T and m variables is used as the input of convolution feature extraction module. A 1-D CNN with a convolution kernel size of $1 \times E$ is applied for each input sequence $x_i = \{x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(T)}\} \in \mathbb{R}^{1 \times T}, i \in (1, m)$ after each convolution operation. Sliding the kernel forward E steps in the time domain ensures that the intervals do not overlap. Through the above operation, the length of the interval is set to be E, and the sequence is divided into $n = T/E$ intervals. Each variable sequence has its own convolution module of CNN_i. Taking the input sequence x_i as an example, the operation process of convolution module $CNN_i(x_i)$ can be expressed as:

$$c_i = CNN_i(x_i) \tag{5}$$

where $c_i = \{c_i^{(1)}, c_i^{(2)}, \ldots, c_i^{(j)}\} \in \mathbb{R}^{n \times J}$ is the feature sequence extracted from the i^{th} variable sequence. To ensure the convolution operation can extract rich features at different levels, J filters are applied to extract J dimensional feature sequences from the variable sequence. $c_i^{(j)} \in \mathbb{R}^{J \times 1}$ is the J-dimensional feature extracted from the j^{th} interval of the i^{th} variable sequence. By convolving the sequence of m input variable, the time interval based J-dimensional feature $C = \{c_1, c_2, \ldots, c_m\} \in \mathbb{R}^{m \times n \times J}$ is determined.

B. Spatial Attention Module

The spatial attention module determines the contribution of each input variable in each interval to the classification based on the spatial attention mechanism. The time interval based J-dimensional feature $C \in \mathbb{R}^{m \times n \times J}$ extracted by convolution feature extraction module is taken as the input of this module. By applying spatial attention mechanism to all input

variables in each interval, the contribution of each input variable is measured in the current interval, and the global spatial feature considering the attention weight of each input variable is obtained.

The $c^{(j)}=\{c_1^{(j)}, c_2^{(j)}, \ldots, c_m^{(j)}\}\in \mathbb{R}^{m\times J}$ extracted by the convolution feature extraction module in Section III-A is sent to the spatial attention module to quantify the spatial attention weight corresponding to each input variable of the current interval [29].

$$e^{(j)}=\zeta(W_s c^{(j)}+b_s) \tag{6}$$

where $e^{(j)}=\{e_1^{(j)}, e_2^{(j)}, \ldots, e_m^{(j)}\}\in \mathbb{R}^{m\times 1}$; ζ is the sigmoid activation function; $W_s\in \mathbb{R}^{J\times 1}$ is the trainable coefficient; and $b_s\in \mathbb{R}^{J\times 1}$ is the bias. Each spatial attention weight is normalized by:

$$\alpha_i^{(j)}=\frac{\exp(e_i^{(j)})}{\sum_{i=1}^{m} e_i^{(j)}} \tag{7}$$

where $\alpha_i^{(j)}\in \mathbb{R}^{1\times 1}$ is the normalized attention weight of each input variable in the current interval and is allowed to meet the probability distribution with the sum of the weights being 1; and exp(·) is the exponential function.

The normalized spatial attention weight is multiplied by the measured value of its corresponding input variable to enhance or weaken the i^{th} input variable and obtain the global spatial feature of the current interval $c'^{(j)}\in \mathbb{R}^{n\times J\times 1}$, considering the weight of each input variable.

$$c'^{(j)}=\sum_{i=1}^{m}\alpha_i^{(j)} c_i^{(j)} \tag{8}$$

The spatial features of each interval are applied to the spatial attention module in turn. $C'=\{c'^{(1)}, c'^{(2)}, \ldots, c'^{(n)}\}\in \mathbb{R}^{J\times n}$ is the global spatial feature sequence of all intervals.

C. Temporal Attention Module

The temporal attention module determines the contribution of each interval to the classification based on the temporal attention mechanism, and then extracts and fuses the temporal features. By applying the temporal attention mechanism to the spatial features of all intervals, the contribution of each interval to the classification is measured, and the global temporal and spatial features considering the attention weight of each interval are obtained.

The extracted global spatial feature sequence of all intervals by the spatial attention module, i.e., C', is applied to the temporal attention module to quantify the temporal attention weight corresponding to each interval.

$$l=\sigma(W_T C'+b_T) \tag{9}$$

where $l=\{l^{(1)}, l^{(2)}, \ldots, l^{(n)}\}\in \mathbb{R}^{n\times 1}$; $W_T\in \mathbb{R}^{J\times 1}$ is the trainable coefficient parameter of the temporal attention module; and $b_T\in \mathbb{R}^{n\times 1}$ is the bias of the temporal attention module. The interval attention weight is normalized by:

$$\beta^{(j)}=\frac{\exp(l^{(j)})}{\sum_{n=1}^{n} l^{(n)}} \tag{10}$$

where $\beta^{(j)}\in \mathbb{R}^{1\times 1}$ is the normalized interval attention weight of each interval and is allowed to meet the probability distribution, with the sum of the weights being 1.

The normalized interval attention weights are then multiplied by the corresponding interval spatial attention to strengthen or weaken the global space feature of the j^{th} interval from:

$$C''=\sum_{j=1}^{n}\beta^{(j)} c'^{(j)} \tag{11}$$

where $C''\in \mathbb{R}^{J\times 1}$ is the global temporal-spatial feature considering the weight of each input variable.

D. Classification Module

The global temporal-spatial feature C'' is used to find the predicted label $\hat{y}^{(T)}$, corresponding to the current sequence X.

$$\hat{y}^{(T)}=Softmax(W_F C''+b_F) \tag{12}$$

where $W_F\in \mathbb{R}^{J\times P}$ is the trainable coefficient parameter of the classification module; $b_F\in \mathbb{R}^{P\times 1}$ is the bias of the classification module; and P is the number of predicted labels .

Assuming θ as the set of coefficients in the CTSAN model, the cross-entropy loss function can be minimized for the training target by iteratively updating the coefficients of each module to finally get the trained proposed CTSAN model in the form of:

$$J(\hat{y}, y, \theta)=-\frac{1}{N}\sum_{k=1}^{N}[y_k \ln(\hat{y}_k)+(1-y_k)\ln(1-\hat{y}_k)] \tag{13}$$

where N is the number of the sample sets; y_k is the true label; and $\hat{y}_k$ is the predicted label.

E. Interpretability

The proposed CTSAN model employs a clear mechanism that assigns weights to spatial variables and time intervals, thereby determining their respective importance. This mechanism facilitates the extraction of critical temporal-spatial features from multi-dimensional SCADA time-series data, and the extracted features are presented as normalized interval attention weights, which directly identify the most significant variables or time intervals for fault classification. The temporal-spatial interpretability of the proposed CTSAN model can be supported from both instance and global perspectives. For instance, the interpretability is assessed by examining the heatmaps of each category, determining the contributions of each input variable within each time interval to the classification. For global interpretability, the interpretability is evaluated by observing the heatmaps and analyzing the trend of the weights of input variables and time intervals.

1) Instance Interpretability

For the input instance sample sets, on the one hand, the spatial attention weight, denoted as $\alpha=\{\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(n)}\}\in \mathbb{R}^{m\times n}$, is output from the spatial attention module. $\alpha^{(j)}=\{\alpha_1^{(j)}, \alpha_2^{(j)}, \ldots, \alpha_m^{(j)}\}\in \mathbb{R}^{1\times m}$ signifies the importance of each input variable within the j^{th} time interval. On the other hand, the normalized interval attention weight, denoted as $\beta=\{\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(n)}\}\in \mathbb{R}^{1\times n}$, is output from the temporal attention module, representing the significance of each time interval. $\chi=\alpha\odot\beta=\{\alpha^{(1)}\beta^{(1)}, \alpha^{(2)}\beta^{(2)}, \ldots, \alpha^{(n)}\beta^{(n)}\}\in \mathbb{R}^{m\times n}$ can be interpreted as the cumulative contribution of each input variable across all time intervals to the classification.

2) Global Interpretability

Global sample sets are acquired from a time period preceding a fault event through sampling. These sample sets are subsequently input into the trained proposed CTSAN model to simulate the global diagnosis process for SCADA data. By examining the predicted fault categories and corresponding temporal-spatial attention weight fluctuations at each time point, the contributions of input variables and time intervals to the fault classification by the proposed CTSAN model are analyzed. For the input global sample sets, the global spatial attention weight $\phi = \sum_{j=1}^{n} \alpha^{(j)} \beta^{(j)} \in \mathbb{R}^{1\times n}$ is defined as the aggregated importance of each input variable across all time intervals, thereby measuring the contribution of each parameter to the fault classification. Concurrently, the temporal attention weight $\beta = \{\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(n)}\} \in \mathbb{R}^{1\times n}$, obtained from the temporal attention module, determines the significance of each time interval, assessing their respective contributions to the fault classification.

IV. Implementation Steps and Performance Indices

A. Flow Chart of Interpretable FD for Offshore WT

Based on the proposed CTSAN model, combined with data preparation and early fault sample set mining, the flow chart of a complete FD process can be obtained in 3 steps, as schematically illustrated in Fig. 4.

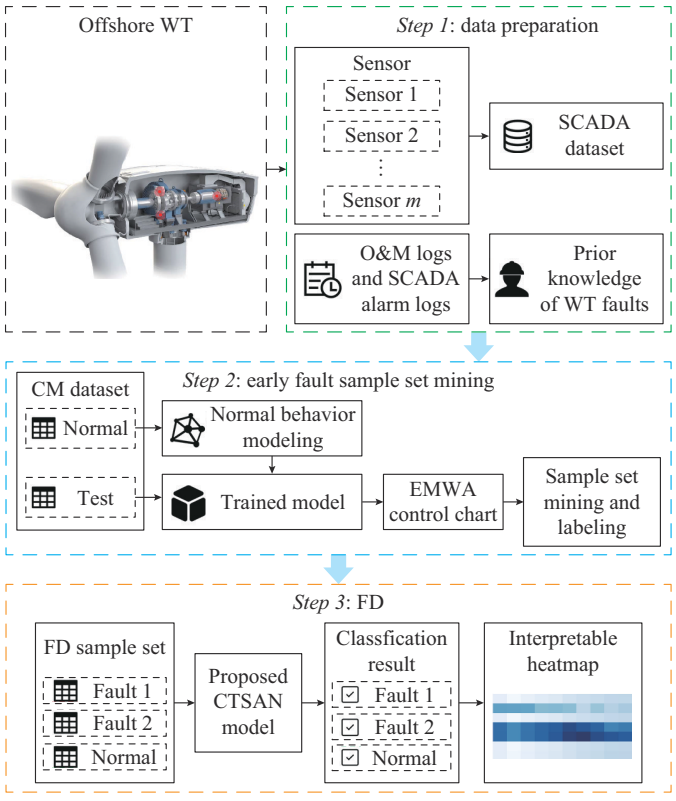

Fig. 4. Flow chart of complete FD process.

Step 1: data preparation. According to the O&M logs and the SCADA alarm logs, the prior knowledge of the WT fault is obtained to study the fault cases. The corresponding dataset is obtained from the SCADA system.

Step 2: early fault sample set mining. The normal operation dataset is obtained for training through data preprocessing, and the NBM process is conducted to obtain the predicted value of the target modelling parameters based on Section II-B. Then, the early fault starting point is determined according to the monitoring process in Section II-C. Finally, according to EWMA control chart, the fault and normal sample set mining and labeling are done, respectively.

Step 3: FD. FD model is trained by the proposed CTSAN model of Section III to yield diagnosis results.

B. FD Evaluation Indicators

To evaluate the accuracy of the proposed CTSAN model, the common performance evaluation indices of accuracy (*ACU*), precision (*P*), recall (*R*) and F1-score (*F*1) are used for the model, while the indices of accuracy, macro-P (*MAP*), macro-R (*MAR*), and macro-F (*MAF*) are used for multi-class models [13]. These indices are mathematically described as:

$$ACU = \frac{TP + TN}{TP + FN + FP + TN} \tag{14}$$

$$\begin{cases} P = \dfrac{TP}{TP + FP} \\ MAP = \dfrac{1}{n}\sum_{i=1}^{n} P_i \end{cases} \tag{15}$$

$$\begin{cases} R = \dfrac{TP}{TP + FN} \\ MAR = \dfrac{1}{n}\sum_{i=1}^{n} R_i \end{cases} \tag{16}$$

$$\begin{cases} F1 = \dfrac{2PR}{P + R} \\ MAF = \dfrac{2 \cdot MAP \cdot MAR}{MAP + MAR} \end{cases} \tag{17}$$

where *TP*, *TN*, *FP*, and *FN* are the true positive, true negative, false positive, and false negative, respectively.

V. Performance Evaluation

The performance of the proposed CTSAN model is evaluated on a 12-month dataset, collected from the SCADA system of Donghai Bridge offshore wind farm in Shanghai, China, in 2018. The wind farm contains 34 WTs. The WT is 3 MW doubly-fed induction generator and the type is SL3000. Each WT is equipped with a SCADA time-series data for condition monitoring and records over 50 readings on different WT components with a sampling frequency of 30 s. Twenty three variables are used for condition monitoring test in this paper, as summarized in Table I [25].

A. Data Preparation

This paper focuses on two types of faults: ① Fault 1 is the fault of high outlet water temperature of generator; and ② Fault 2 is gearbox fault.

Fault 1 and Fault 2 are usually accompanied by the rise of generator shaft temperature and abnormal rise of gearbox oil temperature, respectively.

TABLE I
VARIABLE OF SCADA TIME-SERIES DATA OF WTS

Variable	Notation
Control cabinet temperature	T_{cc}
Generator rotation speed	S_g
Pitch motor torques of WTs 1-3	Tq_{p1}, Tq_{p2}, and Tq_{p3}
Pitch motor temperatures of WTs 1-3	Tq_{pm1}, Tq_{pm2}, and Tq_{pm3}
Pitch angles of WTs 1-3	A_{p1}, A_{p2}, and A_{p3}
Impeller rotational speed	S_i
Active power	P_a
Cabin temperature	T_c
Ambient temperature	T_a
Motor winding temperatures of WTs 1-3	T_{mw1}, T_{mw2}, and T_{mw3}
Shaft temperature of motor non-drive side	T_{mn}
Shaft temperature of motor drive side	T_{md}
Grid current	I
Wind speed	S_w
Gearbox oil temperature	T_{go}

Table II shows the alarm logs of WT-18, in which fault codes 0 and 3987 represent no fault and Fault 1, respectively. The corresponding O&M logs of WT-18 are shown in Table III. The alarm log of WT-20 does not show any fault while Fault 2 is determined by referring to its O&M logs, as shown in Table IV. Through these data, it can be concluded that WT-18 experiences Fault 1 on May 14 while WT-20 experiences Fault 2 around December 12.

TABLE II
ALARM LOGS OF WT-18

Date	Time	Fault code
May 14	20:59:30	0
May 14	21:00:00	3987
May 14	21:00:30	3987
⋮	⋮	⋮
May 15	12:00:00	3987
May 15	12:00:30	0

TABLE III
O&M LOGS OF WT-18

Date	Shutdown reason	Treatment
May 14	Fault 1	Remote restart
May 15	Fault 1	Replace water-cooling fin of generator and add antifreeze

TABLE IV
O&M LOGS OF WT-20

Date	Shutdown reason	Consequence
December 12	Impeller overspeed failure	Wind is too strong to go out to sea
December 14	Abnormal impact inside gearbox	Abnormal sound inside gearbox
December 15	Abnormal impact inside gearbox	Teeth broken in the secondary planetary gear and large ring gear

B. Early Fault Sample Set Mining

1) Data Processing

The PLC status code provided by the SCADA system offers a guide for healthy data extraction [30]. There are totally 12 PLC codes indicating different WT operating modes. Among them, the observations with PLC status code 7 (running) are extracted to construct the healthy dataset.

For WT-18, the SCADA time-series data collected in April are used as training set to train the NBM while the data in May are used as the test set. For WT-20, the data collected in November are used as training set while the data in December are used as the test set. The overview of condition monitoring set is shown in Table V.

TABLE V
OVERVIEW OF CONDITION MONITORING SET

Fault	Data set	Amount	WT No.	Date
Fault 1	Train	86400	WT-18	April 1-April 30
	Test	11520	WT-18	May 14-May 17
Fault 2	Train	86400	WT-20	November 1-November 30
	Test	20160	WT-20	December 5-December 11

2) Fault 1: Early Fault Sample Set Mining

Monitoring the test set of WT-18, it can be observed that when its generator experiences Fault 1, the temperature of motor shaft rises in a short time, so its temperature T_{md} is selected as the target modeling parameter of NBM. Figure 5 shows the fitting situation of T_{md} and control charts of EWMA in the overall range of WT-18.

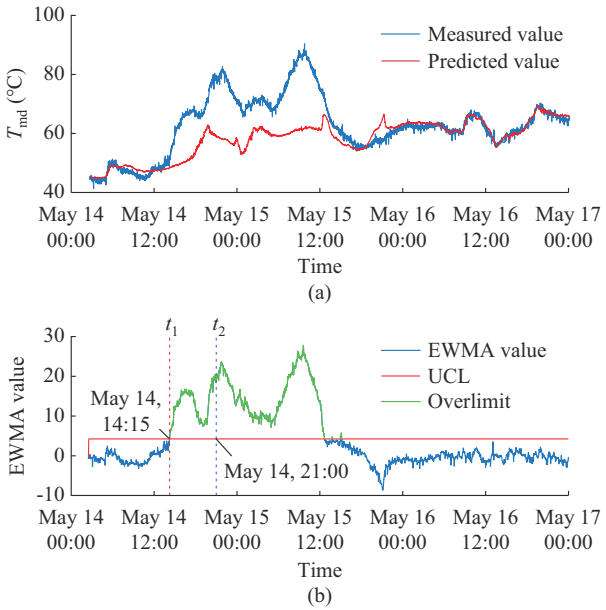

Fig. 5. Monitoring results of WT-18. (a) Fitting situation of T_{md}. (b) Control charts of EWMA.

In this process, t_1 is detected as 6.75 hours before t_2. Figure 5(a) shows that the actual value of T_{md} in the early fault stage is significantly abnormal compared with the predicted value under normal conditions. At the same time, Fig. 5(b) shows that EWMA value in the early fault stage is as obvi-

ous as those in the fault stage, indicating the necessity of early FD. As observed from Fig. 5, the EWMA value exceeds the limit between 14:15 on May 14 and 13:00 on May 15. Therefore, t_1 is assumed as 14:15 on May 14 as the starting point of early fault stage for Fault 1. In addition, the system detects the fault at 21:00 on May 14 when the WT is transitioned into the fault stage. Therefore, t_2 is assumed as 21:00 on May 14 as the starting point of the fault stage.

3) Fault 2: Early Fault Sample Set Mining

Monitoring the test set of WT-20, it can be observed that when the gearbox experiences a fault, the temperature of oil rises in a short time [27], so the temperature T_{go} is selected as the target modeling parameter of NBM. Figure 6 shows the fitting situation of T_{go} and control charts of EWMA in the overall range of WT-20.

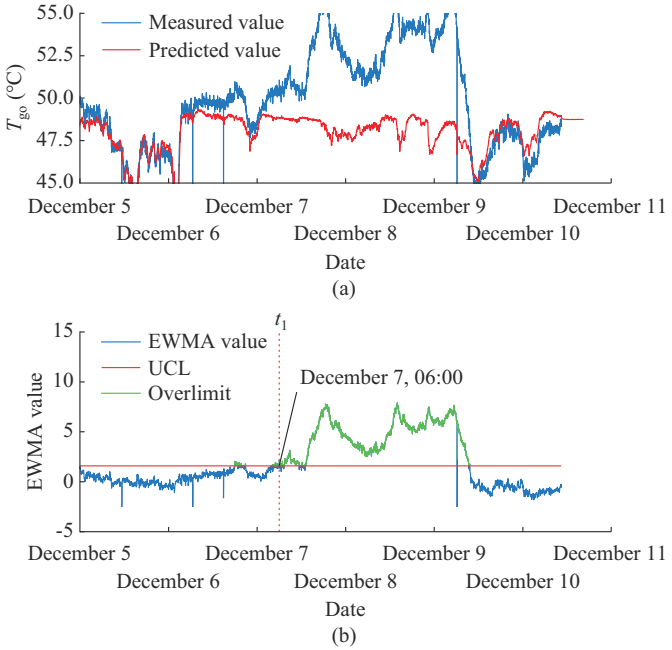

Fig. 6. Monitoring results of WT-20. (a) Fitting situation of T_{go}. (b) Control charts of EWMA.

Figure 6(a) shows that from 06:00 on December 7, T_{go} is significantly abnormal compared with the predicted value under normal working conditions. At the same time, Fig. 6(b) shows that its EWMA value crosses the threshold significantly for 2 days, verifying that the NBM condition monitoring method is effective and can maintain continuous attention to the trend of early fault. However, the fault is not detected by the SCADA data in the whole monitoring interval, which reflects the great limitations of the existing fault monitoring features of SCADA system. The EWMA value exceeds the limit between 06:00 on December 7 and 09:15 on December 9. Therefore, t_1 is assumed as 06:00 on December 7 as the starting point of the early fault stage of Fault 2. Since the system does not detect any faults in the overall monitoring interval, according to the results of WT-18, it is assumed here that the early fault stage will turn into the fault stage 6 hours later. Therefore, t_2 is assumed as the starting point of fault stage at 12:00 on December 7.

C. FD Analysis

1) Sample Set Mining and Labeling

Considering the starting and ending points of fault data assumed above, the normal operational process is 2 weeks before each fault. Set the window length $L=100$, and the normal and fault sample sets are sampled by sliding window method. The sliding step of the normal, early fault, and fault stages are assumed as 50, 5, and 10, respectively.

The overview of FD sample set is shown in Table VI. First, the sample sets of Fault 1 and the normal stage before Fault 1 are selected for single-fault FD verification. Then, the sample sets of Fault 1, Fault 2, the normal stage before Fault 1, and the normal stage before Fault 2 are considered for a multi-fault FD verification. The remaining normal and fault sample sets are combined to a test set under the condition where the sample sets maintain a quantity balance and avoid biased results.

TABLE VI
OVERVIEW OF FD SAMPLE SET

Stage	Sample amount	WT No.	Time period
Fault 1	468	WT-18	14:15 on May 14 to 13:00 on May 15
Fault 2	436	WT-20	06:00 on December 7 to 00:00 on December 8
Normal stage before Fault 1	805	WT-18	00:00 on May 1 to 00:00 on May 13
Normal stage before Fault 2	807	WF20	00:00 on November 22 to 00:00 on December 6

Tables VII and VIII show an overview of the single-fault and multi-fault sample sets, respectively. To reduce influence of different dimensions of input variables on the calculation results, the input set is then normalized and applied to the FD model.

TABLE VII
OVERVIEW OF SINGLE-FAULT SAMPLE SET

Sample set	Type	Sample amount
Train	Fault 1	374
Train	Normal stage before Fault 1	374
Test	Fault 1	94
Test	Normal stage before Fault 1	431

TABLE VIII
OVERVIEW OF MULTI-FAULT SAMPLE SET

Sample set	Type	Sample amount
Train	Fault 1	374
Train	Fault 2	348
Train	Normal stage before Fault 1	722
Train	Normal stage before Fault 2	722
Test	Fault 1	94
Test	Fault 2	88
Test	Normal stage before Fault 1	890
Test	Normal stage before Fault 2	890

2) Model Parameter Setting

In this paper, the proposed CTSAN model is configured with a sequence length of 100, and the convolution module is designed with a kernel size of 1×10 and a stride of 10. Consequently, each sample set is segmented into 10 distinct time intervals. The number of filters is set to be 16. Given the intensive computational requirements for training the proposed CTSAN model, a Windows 10 server equipped with advanced hardware and software specifications is leveraged to facilitate the training process. The hardware configuration of the server includes an NVIDIA RTX 3090 graphics card with 24 GB of memory. CUDA version 11.3 and the PyTorch DL framework is employed. The detailed settings of the proposed CTSAN model is comprehensively outlined in Table IX.

TABLE IX
DETAILED SETTINGS OF PROPOSED CTSAN MODEL

Model setting	Single-fault	Multi-fault
Batch size	64	64
Training epoch	300	300
Learning rate	0.01	0.01
Input sequence length	100	100
Number of convolution feature extraction modules	8	16
Convolution feature extraction module	Kernel size: 1×10 Stride: 10 Filter: 16	Kernel size: 1×10 Stride: 10 Filter: 16
Output dimension of convolution module	64×8×10×16	64×16×10×16
Spatial attention module	Activation: Softmax Filter: 16	Activation: Softmax Filter: 16
Output dimension of spatial attention module	64×8×10	64×16×10
Temporal attention module	Activation: Softmax Filter: 10	Activation: Softmax Filter: 10
Output dimension of temporal attention module	64×8×1	64×16×1
Classification module	Activation: Softmax Output: 64×2	Activation: Softmax Output: 64×3

3) Classification Performance Analysis

To verify the effectiveness and reliability of the proposed CTSAN model, a comparative analysis against different models (i.e., the shallow ML model MLP [6] and DL model STMNN [13]) is conducted and all temporal-spatial features are extracted from the SCADA data. In the experiments, all models are configured with a batch size of 64. The Adam optimizer is employed during the training process, with an initial learning rate of 0.01 and a total of 300 training iterations. The MLP is structured with three fully-connected layers, featuring a hidden layer dimension of 128, and utilizes the Softmax activation function for fault classification. The architecture of the STMNN, as detailed in [13], integrates a deep echo stage network and a multi-scale deep residual CNN with convolution kernels of sizes 1, 3, 5, and 7, facilitating the fusion of temporal-spatial features. To account for the inherent randomness in the model training process and to ensure fairness, all the results are the average performance of three tests.

The performance evaluations are firstly conducted on a single-fault sample set. Figure 7 illustrates the confusion matrices for a single-fault test (Fault 1) based on different models, comparing the performance of two models with the proposed CTSAN model. In the matrix, 0 and 1 represent the normal and fault stages of the sample sets, respectively. The upper left and lower right corners of the matrix display the TP and TN of the sample sets while the lower left and upper right corners display FP and FN of the sample sets. The higher TP and TN illustrate a better classification performance by the FD model. It is evident that MLP has the lowest *TP*, suggesting a risk of overlooking true fault sample sets. Conversely, the proposed CTSAN model and the STMNN yield the highest TP, indicating the successful diagnosis of all sample sets for Fault 1. However, the STMNN model tends to misclassify more normal sample sets, resulting in higher maintenance costs for offshore WTs compared with the proposed CTSAN model. Therefore, the proposed CTSAN model demonstrates superior fault classification accuracy.

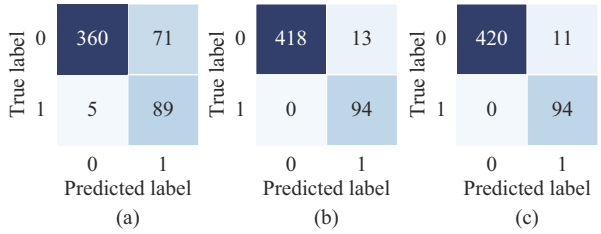

Fig. 7. Confusion matrices for single-fault test based on different models. (a) MLP model. (b) STMNN model. (c) Proposed CTSAN model.

Table X lists the performance evaluation indices of different models for Fault 1 to evaluate the diagnosis accuracy. As can be observed, the proposed CTSAN model has the highest accuracy across all evaluation indices. For instance, compared with the MLP, the proposed CTSAN model has improved the detection *ACU* and *R* by 12% and 13.5%, respectively. This superior performance is attributed to the proposed CTSAN model's comprehensive exploitation of the inherent spatiotemporal features in SCADA data. Furthermore, when compared with the STMNN model, the proposed CTSAN model has improved *F*1 and recall by 0.2% and 0.4%, respectively. This suggests that the temporal-spatial dual attention adopted makes the extraction of temporal-spatial features of the model more comprehensive and effective, resulting in more stable diagnosis.

TABLE X
PERFORMANCE EVALUATION INDICES OF DIFFERENT MODELS FOR FAULT 1

Model	*ACU* (%)	*P* (%)	*R* (%)	*F*1 (%)
MLP	85.5	98.6	83.5	90.5
STMNN	97.5	100.0	97.0	98.5
Proposed CTSAN	97.9	100.0	97.4	98.7

To further explore the performance of the proposed CTSAN model for multi-fault FD, Fig. 8 shows the confusion

matrices for multi-fault test based on the MLP and STMNN, as well as that based on the proposed CTSAN model. In the matrix, 0, 1, and 2 represent the normal state, Fault 1, and Fault 2 stages of the sample sets, respectively. As observed from Fig. 8, the proposed CTSAN model has the highest *TP* and *TN* amongst all models, indicating it has the best classification performance in all tests. Table XI lists the performance evaluation indices of different models for multi-fault FD and shows that the FD accuracy of all three models decreases. Among them, the MLP, constrained by its simplistic architecture and limited feature fitting of the SCADA time-series data, records a mere 79.9% in diagnosis accuracy. Conversely, the proposed CTSAN model has a diagnosis accuracy of 96.7%, outperforming the MLP and STMNN by 16.8% and 0.9%, respectively. It can diagnose Fault 1 and Fault 2 effectively. As for the detection index recall, the proposed CTSAN model achieves an *MAR* value of 98.4%, reflecting the minimal incidence of missed detections and demonstrating its robust capacity to discern between sample sets of fault and normal stages. The proposed CTSAN model uses CNNs to extract key features based on time intervals from the input multi-variate SCADA time-series data, shifting the focus of the model from numerous time points to fewer time intervals. Subsequently, the weight distribution mechanism of temporal-spatial attention is utilized to obtain a more effective feature representation. Based on the temporal-spatial feature representation focused by the attention weights, the classification module can better model the temporal-spatial features among multi-variate SCADA time-series data. Compared with STMNN, which also extracts temporal-spatial features, this further enhances the adaptability of the proposed CTSAN model for multi-fault FD of offshore WTs.

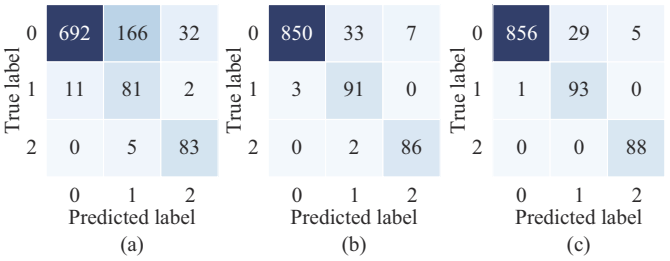

Fig. 8. Confusion matrices for multi-fault test based on different models. (a) MLP model. (b) STMNN model. (c) Proposed CTSAN model.

TABLE XI
PERFORMANCE EVALUATION INDICES OF DIFFERENT MODELS FOR MULTI-FAULT FD

Model	*ACU* (%)	*MAP* (%)	*MAR* (%)	*MAF* (%)
MLP	79.9	67.2	86.1	75.5
STMNN	95.8	88.1	96.7	92.2
Proposed CTSAN	96.7	90.2	98.4	94.1

4) FD Time Performance

For FD, it is crucial to achieve a balance between the accuracy and response time, as a shorter response time gives the offshore WT maintenance system more time to react. Consequently, it is imperative to examine the FD time of different models. Table XII delineates the training time of the proposed CTSAN model and the other two models on both single-fault and multi-fault sample sets, in addition to the single sample set FD time. As indicated in Table XII, the MLP exhibits the shortest training time due to its comparatively simple structure. However, this simplicity also constrains its discriminative capacity of FD for offshore WT, resulting in the accuracy of FD is only at average level. The STMNN characterized by its temporal-spatial feature extraction demonstrates the longest training time in both sample sets, specifically 216.35 s and 246.63 s, thereby incurring a significantly higher computational cost than the proposed CTSAN model. This is attributed to the design of the deep echo stage network of STMNN, which extracts SCADA time-series features at one scale at a time, and the extraction of features at multi-scale must await the completion of the feature extraction of the previous time scale.

TABLE XII
COMPARISON OF MODEL DIAGNOSIS TIME

Fault type	Model	The minimum single epoch training time (s)	Total training time (s)	Single sample set FD time (ms)
Single-fault	MLP	0.16	51.94	1.27
	STMNN	0.67	216.35	2.43
	Proposed CTSAN	0.41	142.61	1.74
Multi-fault	MLP	0.21	68.71	1.32
	STMNN	0.77	246.63	2.51
	Proposed CTSAN	0.49	154.89	1.77

For the practical applications of FD for offshore WTs, the online FD response speed is often more important than the offline training speed. As can be observed from Table XII, the FD time of the three models is less than 3 ms, which to some extent reflects the advantages of data-driven methods, providing more reaction time for offshore WT maintenance personnel. Among them, the proposed CTSAN model has a faster online FD speed compared with the STMNN. In detail, the single sample set FD time of the proposed CTSAN model and STMNN are 1.74 ms and 1.77 ms, respectively, which enables the proposed CTSAN model to deliver diagnosis results in a significantly shorter time, making it particularly suitable for real-time FD for offshore WTs.

D. Interpretability Performance

The interpretability of attention mechanism is manifested in the degree of importance assigned to the features corresponding to attention weights. The features with higher weights typically have a more decisive influence on the diagnosis results. To validate the interpretability of the proposed CTSAN model, the learned weights for both temporal and spatial attention are visualized. The reasonableness of attention weights is also explained by integrating domain-specific knowledge, such as offshore WT fault logs and anomalous behavior of monitored variables.

1) Instance Interpretability Analysis

Figure 9 shows the actual input variable values and interpretable heatmaps of different sample sets, as well as the interpretability analysis of an instance sample set. The vertical and horizontal axes represent m input variables and n intervals, respectively, while the color density illustrates the weight of attention and the contribution of each input variable in each interval to the classification. When the WT is at the normal state stage, the actual input variable values do not fluctuate significantly, as observed from Fig. 9(a). Also, as observed from Fig. 9(c), the weight of input variables such as Tq_{p1}, Tq_{p2}, Tq_{p3}, and T_{mn} at each interval is slightly higher than that of other variables. The proposal has not captured any abnormal conditions from the sample set and thus, it classifies the sample set as normal stage. When the WT shows an early fault trend with T_{go}, T_{md} rises rapidly, and T_{mn} is also high, as observed from Fig. 9(b). Under this condition, as observed from Fig. 9(d), the weights of T_{md} and T_{mn} in the last four intervals are significantly higher than that of other variables, indicating that the proposed CTSAN model detects abnormal conditions of the two variables from the sample set, and based on this, the sample sets are classified (i.e., Fault 1).

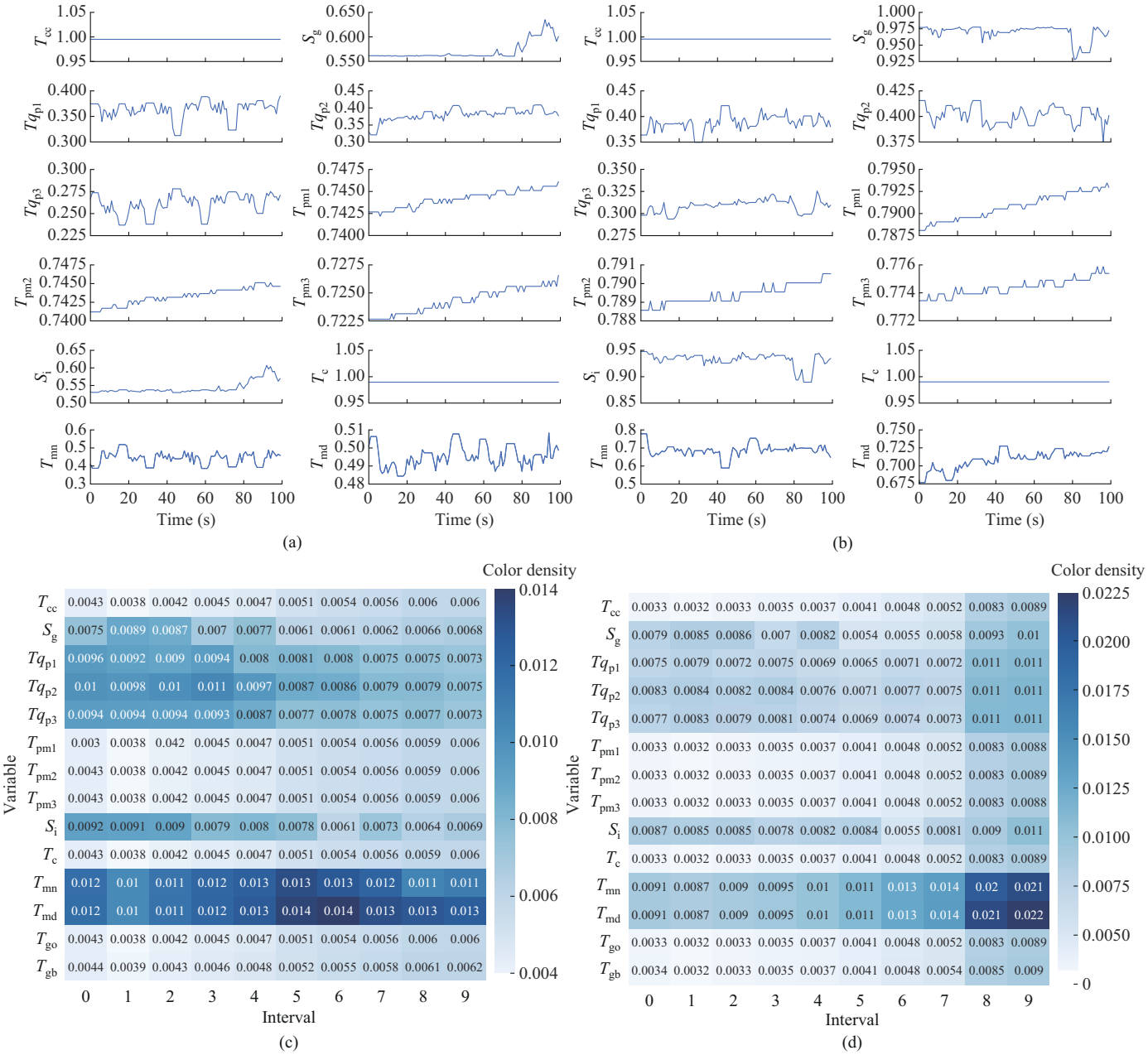

Fig. 9. Actual input variable values and interpretable heatmaps of different sample sets. (a) Normal state stage. (b) Fault 1. (c) Attention weight of normal state stage. (d) Fault 1 attention weight.

2) Global Interpretability Analysis

Figure 10 illustrates the results of global sample sets. Figure 10(a) shows the prediction of the proposed CTSAN model, in which 0 and 1 denote the normal and fault stages, respectively.

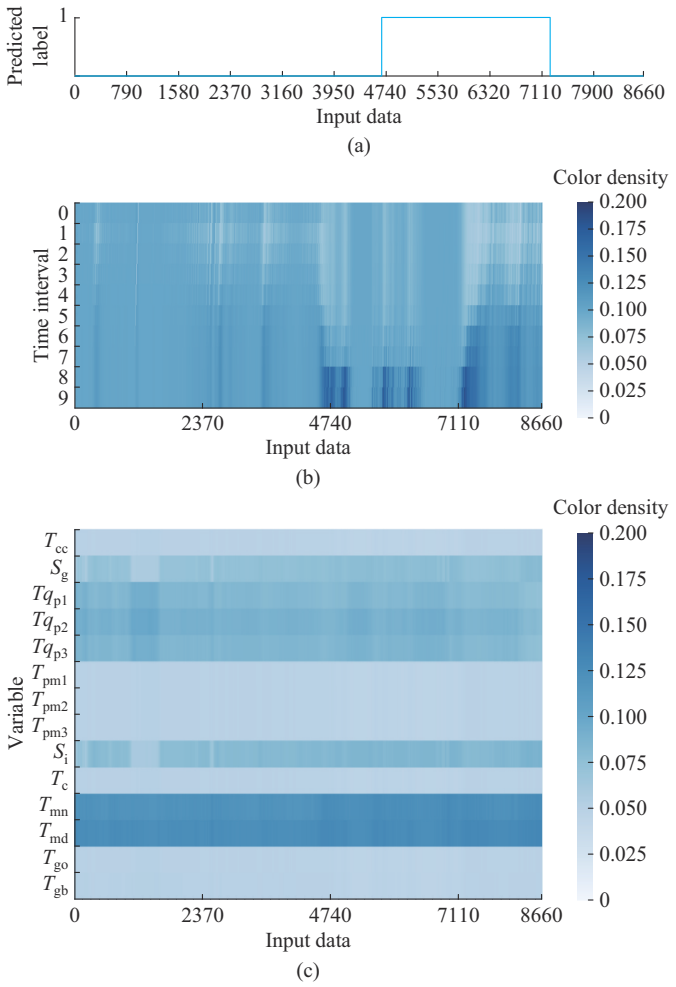

Fig. 10. Results of global sample sets. (a) Prediction of model. (b) Temporal dimension of heatmap. (c) Spatial dimension of heatmap.

Figure 10(b) shows the temporal dimension of heatmap and represents a continuous change process of the temporal attention weight, in which the color density illustrates the attention weight of each interval in the sample sets at different time points and their contribution to the classification.

Figure 10(c) shows the spatial dimension of heatmap and presents a continuous change process of the overall spatial attention weight. When the WT has the trend of an early fault, as observed from Fig. 10(a), the proposed CTSAN model flags a change from normal state stage to fault stage. When the WT shows the trend of an early fault, the temporal attention weight of the end interval rises to 0.2 (much higher than that of other intervals), and temporal attention weight contribution to fault classification is higher. Then, when the WT changes from the early fault stage to the fault stage, the weight of each interval in the sample set and the contribution to the classification are almost the same. After the WT fault is removed, the input variables return to be normal initial values at the end of the sample set interval, and the proposed CTSAN model correctly flags a return to normal stage. Again, the attention weight of these variables at the interval end and the attention weight contribution to the classification is higher. By matching domain knowledge with the heatmaps, it is verified that the proposal successfully extracts the deep temporal-spatial features from the SCADA time-series data of FD for WTs.

VI. Conclusion

An interpretable CTSAN model of FD for offshore WTs is proposed in this paper. By focusing on a real wind farm system, the success of the proposed CTSAN model in detecting the faults is accurately validated. This is demonstrated through two real faults observed in the real system, by comparing the results of the proposed CTSAN model to the actual fault and O&M logs in the real system. This paper illustrates that the proposed CTSAN model can dynamically extract the key deep temporal-spatial features through the available SCADA time-series data, thus improving its FD classification performance. This paper also demonstrates that the proposed CTSAN model is interpretable and shows the extracted deep temporal-spatial features in a human-understandable form, making the evaluation more reliable for practical implementation. The conducted comparisons against two existing models also have numerically proven the superiority of the proposed CTSAN model.

FD for WTs requires the collection of a large amount of data, including mechanical, electrical, and meteorological data, etc. The types of the data are variable, such as thermal imaging from drone inspections, low-frequency SCADA data, and high-frequency mechanical vibration data. Therefore, how to apply the attention mechanism to cross-domain focus on multimodal data, extract features with high discriminability and sensitivity to FD from multiple source features, and integrate the characteristics of algorithms from different domains will be the research focus for the next step in implementing interpretable FD for WTs.

References

[1] F. R. Badal, P. Das, S. K. Sarker *et al*., "A survey on control issues in renewable energy integration and microgrid," *Protection and Control of Modern Power Systems*, vol. 4, no. 1, pp. 1-27, Jan. 2019.

[2] Z. Ren, A. S. Verma, Y. Li *et al*., "Offshore wind turbine operations and maintenance: a state-of-the-art review," *Renewable and Sustainable Energy Reviews*, vol. 144, no. 1, p. 110886, Jul. 2021.

[3] H. Sanchez, T. Escobet, V. Puig *et al*., "Fault diagnosis of an advanced wind turbine benchmark using interval-based ARRs and observers," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3783-3793, Jun. 2015.

[4] Y. Qiu, Y. Feng, J. Sun *et al*., "Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA time-series data," *IET Renewable Power Generation*, vol. 10, no. 5, pp. 661-668, May 2016.

[5] A. Mojallal and S. Lotfifard, "Multi-physics graphical model-based fault detection and isolation in wind turbines," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5599-5612, Nov. 2018.

[6] S. Yang, W. Li, and C. Wang, "The intelligent fault diagnosis of wind turbine gearbox based on artificial neural network," in *Proceedings of 2008 International Conference on Condition Monitoring and Diagnosis*, Beijing, China, Apr. 2008, pp. 1327-1330.

[7] D. Zhang and Z. Qian, "Probability warning for wind turbine gearbox incipient faults based on SCADA time-series data," in *Proceedings of 2017 Chinese Automation Congress (CAC)*, Jinan, China, Oct. 2017, pp. 3684-3688.

[8] M. Mansouri, R. Fezai, M. Trabelsi *et al*., "A novel fault diagnosis of uncertain systems based on interval Gaussian process regression: application to wind energy conversion systems," *IEEE Access*, vol. 8, pp. 219672-219679, Dec. 2020.

[9] J. Lin, H. Shao, X. Zhou *et al*., "Generalized MAML for few-shot cross-domain fault diagnosis of bearing driven by heterogeneous signals," *Expert Systems with Applications*, vol. 230, p. 120696, Nov.

2023.
[10] Z. Xu, M. Bashir, Y. Yang *et al*., "Multisensory collaborative damage diagnosis of a 10 MW floating offshore wind turbine tendons using multi-scale convolutional neural network with attention mechanism," *Renewable Energy*, vol. 199, pp. 21-34, Nov. 2022.
[11] C. Zhang, D. Hu, and T. Yang, "Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost," *Reliability Engineering & System Safety*, vol. 222, p. 108445, Jun. 2022.
[12] Y. Pang, Q. He, G. Jiang *et al*., "Spatio-temporal fusion neural network for multi-class fault diagnosis of wind turbines based on SCADA time-series data," *Renewable Energy*, vol. 161, pp. 510-524, Dec. 2020.
[13] Q. He, Y. Pang, G. Jiang *et al*., "A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA time-series data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6875-6884, Oct. 2021.
[14] X. Qian, T. Sun, Y. Zhang *et al*., "Wind turbine fault detection based on spatial-temporal feature and neighbor operation state," *Renewable Energy*, vol. 219, p. 119419, Dec. 2023.
[15] H. Habibi, I. Howard, and S. Simani, "Reliability improvement of wind turbine power generation using model-based fault detection and fault tolerant control: a review," *Renewable Energy*, vol. 135, p. 877-896, May 2019.
[16] N. Freire and A. Cardoso, "Fault detection and condition monitoring of PMSGs in offshore wind turbines," *Machines*, vol. 9, p. 260, Oct. 2021.
[17] P. Bangalore and M. Patriksson, "Analysis of SCADA time-series data for early fault detection, with application to the maintenance management of wind turbines," *Renewable Energy*, vol. 115, pp. 521-532, Jan. 2018.
[18] S. Wei, X. Zhang, F. Yang *et al*., "Early-fault warning and diagnosis of offshore wind DFIG based on GRA-LSTM-Stacking model," *Proceedings of the CSEE*, vol. 41, pp. 2373-2383, Apr. 2021.
[19] T. Sun, G. Yu, M. Gao *et al*., "Fault diagnosis methods based on machine learning and its applications for wind turbines: a review," *IEEE Access*, vol. 9, pp. 147481-147511, Oct. 2021.
[20] R. Dwivedi, D. Dave, H. Naik *et al*., "Explainable AI (XAI): core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1-33, Sept. 2023.
[21] S. Chaudhari, V. Mithal, G. Polatkan *et al*., "An attentive survey of attention models," *ACM Transactions on Intelligent Systems and Technology*, vol. 12, no. 5, pp. 1-32, Oct. 2021.
[22] Y. Xiao, H. Shao, M. Feng *et al*., "Towards trustworthy rotating machinery fault diagnosis via attention uncertainty in transformer," *Journal of Manufacturing Systems*, vol. 70, pp. 186-201, Oct. 2023.
[23] Y. Li, Z. Zhou, C. Sun *et al*., "Variational attention-based interpretable transformer network for rotary machine fault diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 5, pp. 6180-6193, May 2024.
[24] H. Zhao, H. Liu, H. Liu *et al*., "Condition monitoring and fault diagnosis of wind turbine generator based on stacked autoencoder network," *Automation of Electric Power Systems*, vol. 42, no. 11, pp. 102-108, May 2018.
[25] X. Su, Y. Shan, C. Li *et al*., "Spatial-temporal attention and GRU based interpretable condition monitoring of offshore wind turbine gearboxes," *IET Renewable Power Generation*, vol. 16, no. 2, pp. 402-415, Feb. 2022.
[26] Z. Kong, B. Tang, L. Deng *et al*., "Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA time-series data by convolutional neural networks and gated recurrent units," *Renewable Energy*, vol. 146, pp. 760-768, Jul. 2020.
[27] L. Wang, Z. Zhang, H. Long *et al*., "Wind turbine gearbox failure identification with deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1360-1368, Jun. 2017.
[28] D. Zhang, L. Qian, B. Mao *et al*., "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020-21031, Apr. 2018.
[29] M. T. Luong, H. Pham, and C. D. Manning. (2015, Aug.). Effective approaches to attention-based neural machine translation. [Online]. Avaulable: https://aclanthology.org/D15-1166/
[30] L. Wei, Z. Qian, and H. Zareipour, "Wind turbine pitch system condition monitoring and fault detection based on optimized relevance vector machine regression," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2326-2336, Oct. 2020.

Xiangjing Su received the Ph.D. degree in electrical engineering from Curtin University, Perth, Australia, in 2015. He was a Visiting Fellow with the University of New South Wales, Sydney, Australia, between 2018 and 2020. He is currently working as the Guangming Scholar at Shanghai University of Electric Power, Shanghai, China. His research interests include distribution system planning and operation and offshore wind farm operation and maintenance.

Chao Deng received the B.E. degree in electrical engineering from Hebei Agricultural University, Baoding, China, in 2021, and the M.S. degree in electrical engineering from Shanghai University of Electric Power, Shanghai, China, in 2024. His research interests include offshore wind turbine fault diagnosis, power cyber-physical system security, and machine learning.

Yanhao Shan received the B.E. and M.S. degrees in electrical engineering from Shanghai University of Electric Power, Shanghai, China, in 2018 and 2021, respectively. He is currently working as an Engineer at Yantai Power Supply Company, State Grid Shandong Electric Power Co., Ltd., Yantai, China. His research interests include condition monitoring of offshore wind power.

Farhad Shahnia received the Ph.D. degree in electrical engineering from the Queensland University of Technology, Brisbane, Australia, in 2012. He is currently an Associate Professor with Murdoch University, Perth, Australia. His research interests include application and control of power electronic converters in distribution system and microgrid.

Yang Fu received the Ph.D. degree in electrical engineering from Shanghai University, Shanghai, China, in 2007. He is currently a Professor and the Vice President of Shanghai University of Electric Power, Shanghai, China. His research interests include offshore wind power planning and operation and new power system analysis and optimization.

Zhaoyang Dong received the Ph.D. degree from University of Sydney, Sydney, Australia, in 1999. He is currently a Professor of power engineering with Nanyang Technological University, Singapore, and a Singapore Power Group endowed Professor of power engineering. His previous roles include a SHARP Professor of energy systems, the inaugural Director of University of New South Wales Digital Grid Future Institute and University of New South Wales, Sydney, Australia, the Director of the ARC Research Hub for Integrated Energy Storage Systems, the Head of the School of Electrical and Information Engineering, University of Sydney, and the Ausgrid Chair Professor and the Director of the Ausgrid Centre for Intelligent Electricity Networks. His research interests include power system planning and stability, smart grid, smart city, renewable energy system, and energy market.